

# Question 1

Begin your response to QUESTION 1 on this page.

## STATISTICS

### SECTION II

Total Time—1 hour and 30 minutes

6 Questions

### SECTION II, Part A

Suggested Time—1 hour and 5 minutes

5 Questions

**Directions:** Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

1. A large exercise center has several thousand members from age 18 to 55 years and several thousand members age 56 and older. The manager of the center is considering offering online fitness classes. The manager is investigating whether members' opinions of taking online fitness classes differ by age. The manager selected a random sample of 170 exercise center members ages 18 to 55 years and a second random sample of 230 exercise center members ages 56 years and older. Each sampled member was asked whether they would be interested in taking online fitness classes.

The manager found that 51 of the 170 sampled members ages 18 to 55 years and that 79 of the 230 sampled members ages 56 years and older said they would be interested in taking online fitness classes.

At a significance level of  $\alpha = 0.05$ , do the data provide convincing statistical evidence of a difference in the proportion of all exercise center members ages 18 to 55 years who would be interested in taking online fitness classes and the proportion of all exercise center members ages 56 years and older who would be interested in taking online fitness classes? Complete the appropriate inference procedure to justify your response.

	18-55	56+
$n$	170	230
$\hat{p}$	$\frac{51}{170} = 0.3$	$\frac{79}{230} = 0.3434$

2-Sample z test

$$z^* = \frac{0.3434 - 0.3}{\sqrt{\frac{(0.3)(1-0.3)}{170} + \frac{(0.3434)(1-0.3434)}{230}}} = 1.8329$$

$$p_c = \frac{51 + 79}{170 + 230} = 0.342105$$

$$p\text{-value} = (0.033408) \times 2 = 0.066817$$

$\therefore p\text{-value} > \alpha$

there is convincing statistical evidence of difference in the age group 18-55 by 2.56

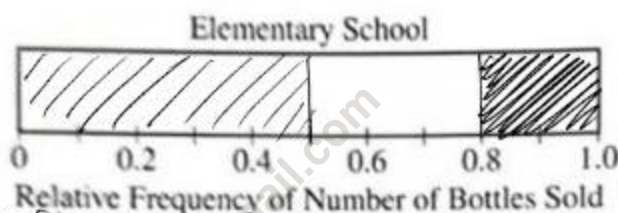
2. A local elementary school decided to sell bottles printed with the school district's logo as a fund-raiser. The students in the elementary school were asked to sell bottles in three different sizes (small, medium, and large). The relative frequencies of the number of bottles sold for each size by the elementary school were 0.5 for small bottles, 0.3 for medium bottles, and 0.2 for large bottles.

A local middle school also decided to sell bottles as a fund-raiser, using the same three sizes (small, medium, and large). The middle school students sold three times the number of bottles that the elementary school students did. For the middle school students, the proportion of bottles sold was equal for all three sizes.

- (a) Complete the segmented bar graphs representing the relative frequencies of the number of bottles sold for each size by students at each school.

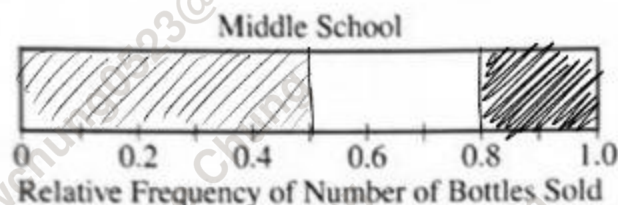
Small: 0.5  
Medium: 0.3  
Large: 0.2

proportion



Key

- Small bottles
- Medium bottles
- Large bottles



- (b) An administrator at the elementary school concluded that the elementary school students sold more small bottles than the middle school students did. Is the elementary school administrator's conclusion correct? Explain your response.

No, middle school students sold more small bottles than the elementary school.

Total # of bottles sold by middle is 3 times larger, while the proportion of each bottle stayed the same.

This means that middle school sold 3 times more small, medium and large bottles than the elementary.



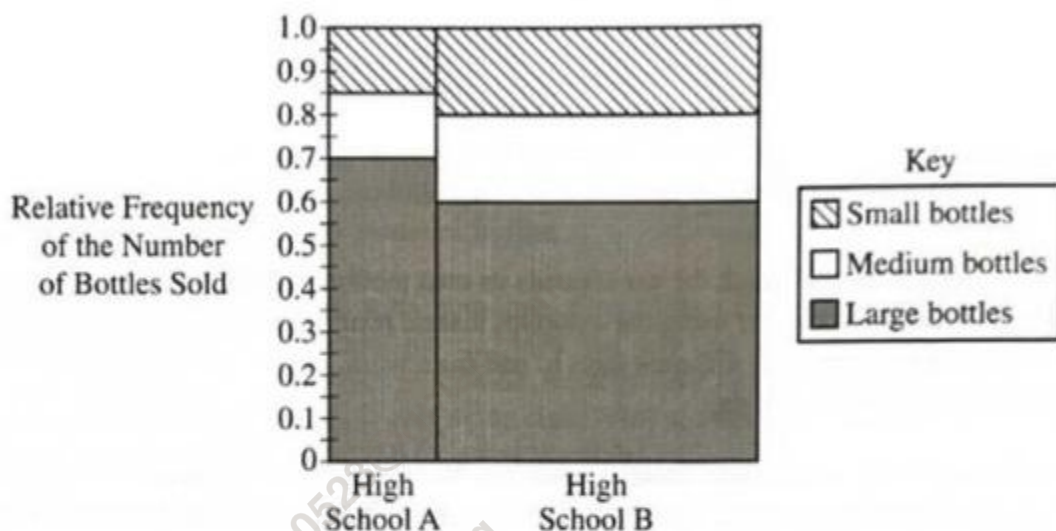
## Question 2

Continue your response to **QUESTION 2** on this page.

Two high schools are also selling the bottles and are competing to see which one sold more large bottles.

(c) A mosaic plot for the distribution of the number of bottles sold by each of the high schools is shown here.

**Distribution of the Number of Bottles Sold by High School**



(i) Which of the two high schools sold a greater proportion of large bottles? Justify your answer.

High School A.  
 Along the y-axis A has higher height  
 larger frequency  
 A = 0.7 while B = 0.6.

(ii) Which of the two high schools sold a greater number of large bottles? Justify your answer.

High School B.  
 horizontal length representing the  
 relative number of bottles sold  
 High school B has larger horizontal  
 length than A, even though its height (frequency)  
 is smaller, the overall area is greater.  
 ∴ "number" of bottles is higher in high school B.

### Question 3

Begin your response to **QUESTION 3** on this page.

31. A car maker produces four different models of cars: A, B, C, and D. A group of researchers is investigating which model of car has the longest distance traveled per gallon of gas (mileage). Higher mileage is considered better than lower mileage. The researchers will conduct a study in which they contact several owners of each model of car and ask them to estimate their mileage.

(a) Is this an observational study or an experiment? Justify your answer in context.

observational

they are not selecting people randomly to assign a type of car. Rather, they collect the data that are already present.

It contacted the owners and got data from them.

Model D has an autopilot feature, in which the car controls its own motion with human supervision. James owns a Model D car and will investigate whether using the autopilot feature results in higher mileage than not using the autopilot. James will drive his car on 30 different days to and from work, using the same route at the same time each day. James will record the mileage each day.

- (b) James will use a completely randomized design to conduct his investigation. Describe an appropriate method James could use to randomly assign the two treatments, driving using the autopilot feature and driving without using the autopilot feature, to 15 days each.

- (c) After the investigation was completed, James verified that the conditions for inference were met and conducted a hypothesis test. He discovered the mean mileage when using the autopilot feature was significantly higher than the mean mileage when not using the autopilot feature.

James is a member of a Model D club with thousands of members who all drive Model D cars. He will give a presentation at a Model D club members' meeting later this year and would like to state that the results of his hypothesis test apply to all Model D cars in his club. Another member of the club who is a statistician tells James his findings do not apply to all Model D cars in the club. What change would James need to make to his original study to be able to generalize to all Model D cars in the club?

wchung0523@gmail.com  
Wonjae Chung

22 April 2025

AlISATpapers.com

Single-person usage only



Begin your response to **QUESTION 4** on this page.

4. In an online game, players move through a virtual world collecting geodes, a type of hollow rock. When broken open, these geodes contain crystals of different colors that are useful in the game. A red crystal is the most useful crystal in the game. The color of the crystal in each geode is independent and the probability that a geode contains a red crystal is 0.08.

(a) Sarah, a player, will collect and open geodes until a red crystal is found.

- (i) Calculate the mean of the distribution of the number of geodes Sarah will open until a red crystal is found. Show your work.

$$E(X) = \frac{1}{p} = \frac{1}{0.08} = 12.5$$

- (ii) Calculate the standard deviation of the distribution of the number of geodes Sarah will open until a red crystal is found. Show your work.

\*Geometric Distribution

$$V(X) = \sigma^2(X) = \frac{1-p}{p^2}$$

$$\sigma(X) = \sqrt{\frac{(1-p)}{p^2}}$$

$$= \sqrt{\frac{1-0.08}{0.08^2}}$$

$$= 11.98957 \approx 11.99$$

- (b) Another player, Conrad, decides to play the game and will stop opening geodes after finding a red crystal or when 4 geodes have been opened, whichever comes first. Let  $Y$  = the number of geodes Conrad will open. The table shows the partially completed probability distribution for the random variable  $Y$ .

Number of geodes Conrad will open, $y$	1	2	3	4
Probability, $P(Y = y)$	0.08	0.0736	0.067712	0.778688

$$\begin{array}{r} 0.08 \\ (0.92)^1(0.08) \\ (0.92)^2(0.08) \\ (0.92)^3(0.08) \\ \hline 0.778688 \end{array}$$

(i) Calculate  $P(Y = 3)$ . Show your work.

(ii) Calculate  $P(Y = 4)$ . Show your work.

(c) Consider the table and your results from part (b).

(i) Calculate the mean of the distribution of the number of geodes Conrad will open. Show your work.

$$E(Y) = 1(0.09) + 2(0.0736) + 3(0.067712) + 4(0.719698)$$

mean  $\Rightarrow$  3.545

(ii) Interpret the mean of the distribution of the number of geodes Conrad will open, which was calculated in part (c-i).

On average, Conrad will open geodes 3.545 times before opening the real crystal or opened the max times (4) without finding the Real crystal.



5. Baseball cards are trading cards that feature data on a player's performance in baseball games. Michelle is at a national baseball card collector's convention with approximately 20,000 attendees. She notices that some collectors have both regular cards, which are easily obtained, and rare cards, which are harder to obtain. Michelle believes that there is a relationship between the number of months a collector has been collecting baseball cards and whether the majority of the cards (cards appearing more often) in their collection are regular or rare. She obtains information from a random sample of 500 baseball card collectors at the convention and records how many full months they have been collecting baseball cards and whether the majority of the cards in their card collection are regular or rare. Her results are displayed in a two-way table.

**Majority Type of Baseball Cards and Months of Collecting Baseball Cards**

	Fewer Than 6 Months	6 - 10 Months	11 - 15 Months	16 - 20 Months	21 or More Months	Total
Has a Majority of Regular Baseball Cards	80	84	71	76	112	423
Has a Majority of Rare Baseball Cards	11	16	9	6	35	77
Total	91	100	80	82	147	500

- (a) If one collector from the sample is selected at random, what is the probability that the collector has been collecting baseball cards for 11 or more months and has a majority of regular baseball cards? Show your work.

$$P(x) = \frac{71 + 76 + 112}{500} = \frac{259}{500} = 0.518$$

- (b) Given that a randomly selected collector from the sample has been collecting baseball cards for fewer than 6 months, what is the probability the collector has a majority of regular baseball cards? Show your work.

$$P(\text{regular} \mid \text{fewer than 6}) = \frac{P(\text{regular} \cap \text{fewer than 6})}{P(\text{fewer than 6})} = \frac{80}{91}$$

$$\approx 0.87912$$

$$\approx 0.88$$



Continue your response to **QUESTION 5** on this page.

- (c) Michelle believes there is a relationship between the number of months spent collecting baseball cards and which type of card is the majority in the collection (regular or rare).
- (i) Name the hypothesis test Michelle should use to investigate her belief. Do not perform the hypothesis test.

chi-test

- (ii) State the appropriate null and alternative hypotheses for the hypothesis test you identified in (c-i). Do not perform the hypothesis test.

$H_0$ : there is no relationship b/w # months & types of card  
 $H_a$ : there is relationship

- (d) After completing the hypothesis test described in part (c), Michelle obtains a  $p$ -value of 0.0075. Assuming the conditions for inference are met, what conclusion should Michelle make about her belief? Justify your response.

Since the  $p$ -value 0.0075 is smaller than any common  $\alpha$  value such as 0.05, there is convincing evidence that there is a relationship between # months & types of card

## Question 6

Begin your response to **QUESTION 6** on this page.

### SECTION II, Part B

Suggested Time—25 minutes

#### 1 Question

**Directions:** Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. A company sells a certain type of whistle. The price of the whistle varies from store to store. Julio, a statistician at the company, wants to estimate the mean price, in dollars (\$), of this type of whistle at all stores that sell the whistle.

(a) (i) Identify the appropriate **inference procedure** for Julio to use.

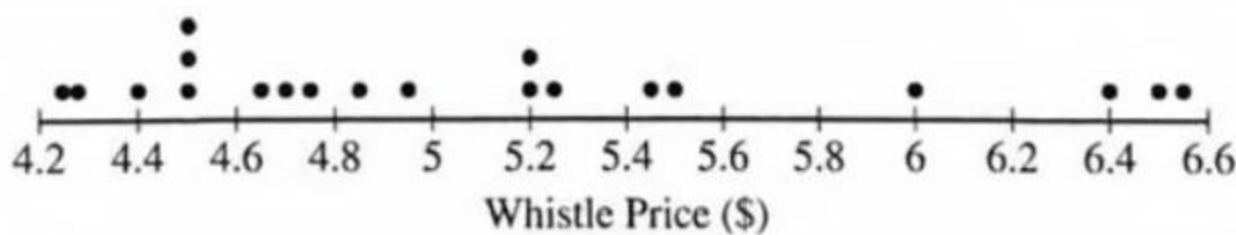
survey  
One Sample t-interval

(ii) Describe the parameter for the inference procedure you identified in part (a-i) in context.



Julio called the managers of 20 randomly selected stores that sell the whistle and recorded the price of the whistle at each store. Following is a dotplot of Julio's data.

Price of the Whistle at 20 Stores





The summary statistics for Julio's data are shown in the following table.

Summary Statistics for Julio's Data

Sample Size	Mean	Standard Deviation	Minimum	$Q_1$	Median	$Q_3$	Maximum
20	5.12	0.743	4.25	4.51	4.885	5.475	6.58

$$IQR: 1.225$$

(b) Julio wants to examine some characteristics of the distribution of the sample of whistle prices.

(i) Describe the shape of the distribution of the sample of whistle prices. Justify your response using appropriate values from the summary statistics table.

Skewed to right  $Q_1 - \text{Median} : 0.375$   
 $\text{Median} - Q_3 : 0.59$

25th are less variability than 75th.

there is gap

Range: 2.33

IQR: 1.225

(ii) Using the  $1.5 \times IQR$  rule, determine whether there are any outliers in the sample of whistle prices. Justify your response.

If a value is outside of range  $(Q_1 - 1.5IQR, Q_3 + 1.5IQR)$ , it is considered as an outlier.

$$(4.51 - 1.225 \times 1.5, 5.475 + 1.225 \times 1.5) \\ = (2.6725, 7.3125)$$

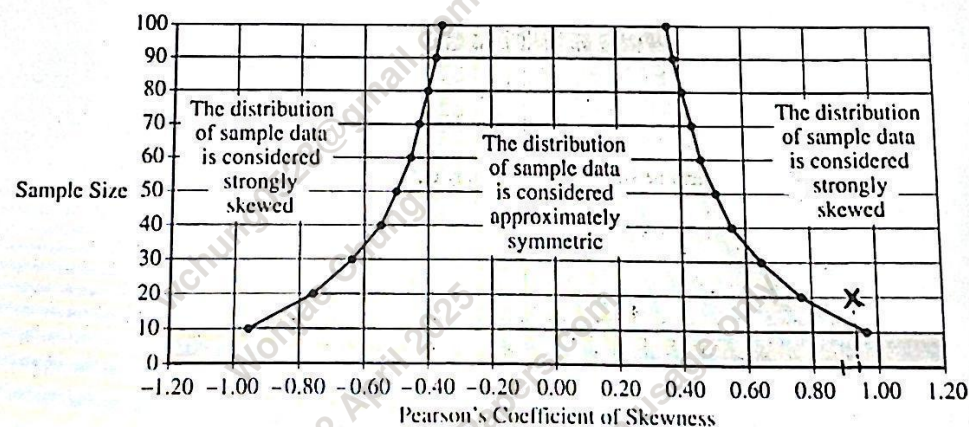
An outlier is either smaller than 2.6725 or greater than 7.3125. Since the minimum & maximum are both within the range, there is no outlier.

(c) (i) Calculate Pearson's coefficient of skewness for Julio's sample of 20 whistle prices. Show your work.

$$\frac{3(\bar{x} - m)}{s} = \frac{3(5.12 - 4.885)}{0.743} = 0.94885.$$

The following graph shows conclusions that can be made about the shape of the distribution of sample data based on Pearson's coefficient of skewness and sample size.

Conclusion from Pearson's Coefficient of Skewness



(ii) Indicate the value of the Pearson's coefficient of skewness you calculated in part (c-i) for the appropriate sample size by marking it with an "X" on the preceding graph.

\* Sample size: 20  
Coefficient: 0.94885

Unauthorized copying or reuse of this page is illegal.

Page 14

GO ON TO THE NEXT PAGE

Use a pencil or a pen with black or dark blue ink. Do NOT write your name. Do NOT write outside the box.

0138413



Continue your response to QUESTION 6 on this page.

Consider your work in part (c).

(d) What should you conclude about the shape of the distribution of the sample of whistle prices? Justify your response.

Since the "x" mark on the graph falls under unshaded region, the shape of the distribution of the sample is "strongly skewed".

Julio's inference procedure in part (a-i) needs one of the following requirements to be satisfied to verify the normality condition.

- The sample size is greater than or equal to 30.
- If the sample size is less than 30, the distribution of the sample data is not strongly skewed and does not have outliers.

(ii) Using your response to (d-i) and the preceding requirements, is the normality condition satisfied for Julio's data? Explain your response.

Since the sample distribution is strongly skewed and the sample size is smaller than 30, the normality condition is not satisfied.